

Modeling of a Demethanizer Tower Using Statistical Tools

Fernando D'Almeida and Carlos Pires

Chemical Reaction Engineering Laboratory, Polytechnic School, UFBA (Federal University of Bahia), Salvador-BA 40210-630, Brazil

Abstract: Monitoring of industrial plant performance and detection on flaws is important to the successful operation on industrial production units. Malfunctioning equipment can greatly impact plant performance by reducing the efficiency and increasing the production cost. Phenomenological equations cannot properly describe industrial processes. Thus, it is necessary to develop new equations for model industrial operations. The purpose of this study is to develop an empirical model for industrial demethanizer tower which is malfunctioning due to an error in the design in one of its plates. A nonlinear statistical model was designed to predict the pressure variation in the column, and consequently, the flooding conditions. This model was validated using industrial data to predict the maximum loads in the column.

Key words: Demethanizer tower, planning experiments, industrial data, empirical model.

1. Introduction

Performance monitoring of equipment and detection of flaws are the key to the successful operation for industrial units. The performance can be monitored by comparing real data with the data predicted by a model [1]. However, phenomenological models are not always able to predict the behavior of malfunctioning pieces of equipment [2, 3]. The use of the multivariate statistical modeling has great potential for industrial application, especially for modeling equipment with physical irregularities [4, 5]. For example, multivariate statistical modeling can be applied for analyzing fallen or obstructed plates in distillation columns, catalyst poisoning, or even the contamination of a stream by an industrial compound.

In early 1990, several scholars began to investigate the use of statistical models for representing phenomena that occur in industrial plants [6]. A formative study on multivariate modeling that laid the groundwork for many statistical models and became a reference for future studies [1]. Later on, the

development of routine processes for computational data collection enabled much greater access to operational data, thus leading to more rigorous statistical treatments [5]. Moreover, Zullo [7] and Kourti et al. [8] proposed the use of statistical modeling for monitoring industrial plants. By the end of the 1990, the first studies using statistical modeling for analyzing malfunctioning equipment began to emerge [4, 5, 9].

Rapid detection and diagnosis of malfunctioning equipment is important because it reduces costs and environmental impacts [5, 9, 10]. Multivariate statistical methods offer the abilities to handle large amounts of data, correlate variables, and measure errors [8]. These methods make it possible to generate models that reproduce the plant operations, such that any deviations can be easily recognized. Because the variables which contribute to the processes are known, it is much easier to diagnosis any problems [2, 8].

2. Material and Methods

2.1 Demethanizer Tower

The demethanizer tower is a typical example of a

Corresponding author: Carlos Pires, Ph.D., research field: process. E-mail: cap@ufba.br.

column with a divided load [11]. A typical configuration for such a column is shown in Fig. 1.

The demethanizer tower is fed by four streams inserted in different plates. The streams arrive from the condensation units, each having different temperatures. Condensation is achieved using refrigeration cycles and product/feed heat exchangers.

Ethylene is the most desirable product from this process, and it is present in all of the liquid condensates. The ethylene which is not condensed leaves the system with methane in the combustible gas stream. The loss of ethylene through this stream should be minimized. Part of the gas that exits the tower is cooled and continuously fed. A typical feed is composed of H_2 , CO , CH_4 , C_2H_2 , C_2H_4 , C_2H_6 , C_3H_4 , C_3H_6 , C_3H_8 , C_4H_6 , C_4H_8 , C_4H_{10} , and C_5+ at a temperature of $15\text{ }^\circ\text{C}$, a pressure of 44 bar, and a flow rate of $4,870\text{ kmol}\cdot\text{h}^{-1}$.

Three product streams (the first two with low ethylene content) are obtained from the plant [11]. The stream used to produce fuel gas is composed of 93.2% methane and 6.5% hydrogen. The stream used to produce 99.9% H_2 is composed of 80.9% hydrogen and 18.7% methane. The stream at the bottom is composed of 68.0% ethylene and other heavier hydrocarbons.

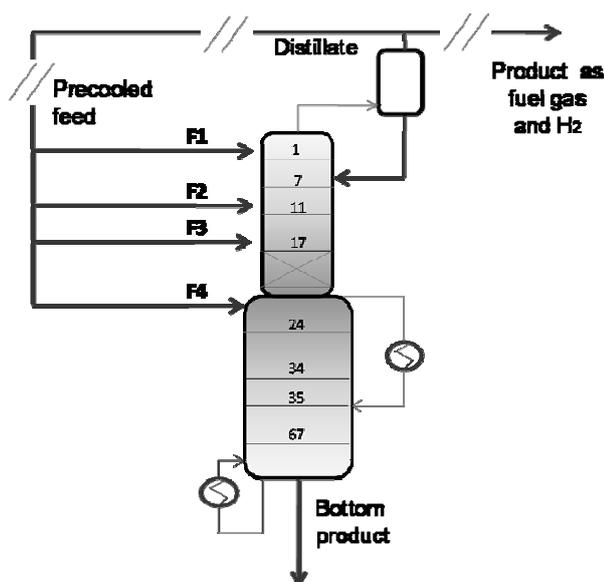


Fig. 1 Demethanizer tower.

2.2 Planning Matrix

Statistical design of experiments was used to create equations to represent the column head loss and predict the maximum operation load. The traditional methodology generates a matrix of tests to determine the responses from experimental actions. This methodology is easy to implement in the laboratory but is very difficult to use in an industrial system. The handling of operational variables in an industrial unit is limited, and thus, the variables are generally extrapolated from values outside of the working range, for example, using out-of-specification products that contribute to economic loss.

The solution for the industrial scenario involves using past operational data which are archived in the unit's database. Data from over 10,000 operational variables and their respective responses (independent and dependent variables) were used to create the planning matrix. The challenge in this procedure was to find the values of the operational variables and their respective responses that corresponded to each test in the planning matrix.

Fig. 2 shows a simplified diagram of the stages which were followed to obtain the values for the operational variables and their respective responses.

When planning any experiment, the first step is to determine the factors (operational or independent variables) and the response of interest (dependent variable) in the process.

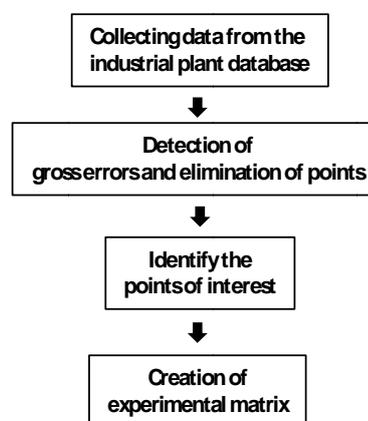


Fig. 2 Experimental steps in the design of experiments process.

Six independent variables were chosen which could influence the phenomenon in the demethanizer tower: the four feed streams in the tower, the reflux flow rate, and the thermal load. The range for each independent variable (with extremes coded from -1 to +1) was defined according to the operational experiment. The extreme values were used to create the planning matrix. A program was built to identify the operational conditions for each test and the responses within the 10,000 archived points. For each set of points (operational conditions), the program scanned and recognized the archived points with an error of $\pm 2\%$ in addition to determining the frequency. The mean values of the operational conditions were then determined, and their responses were inserted into the planning matrix.

The combined values of the six operational variables defined in the planning matrix were not always part of the archived data. Despite using a universe of 10,000 points, some combinations were not identified. In this case, the mean of the combination with the least number of variables was used, which led to a distortion of the results.

To validate the existence of the points, a variation of $\pm 3\%$ with respect to the searched values was initially considered.

2.3 Empirical Equation

An empirical equation to represent the process was built using planning matrix (X) and its respective

responses (Y). The equation coefficients were determined by least squares regression of the parameters.

$$\text{Coefficients} = (X^T X)^{-1} (X^T Y) \quad (1)$$

After determining the empirical equation for the phenomenon, the model was validated using the real data from the plant.

3. Results

3.1 Selection of Variables

The 10,000 archived points were evaluated. The results are shown in Table 1.

The values of the operational variables included in the planning matrix were represented the lower levels of the operational variables (-1), the central points (0), and the higher levels (+1). The central points were defined as the means of the maximum and minimum values of operational variables. The values of the (-1) and (+1) levels were defined as ± 1.2 times the standard deviation added to the central points. Amplitudes of ± 1.2 times the standard deviations were responsible for generating the greatest necessary combination needed to create the planning matrix.

The planning matrix generated with six variables included approximately 50% of the points. To fill the gaps, the mean of the combinations of five variables was used. This shows the difficulty of finding sets of specific values for operational variables, even when

Table 1 Operational variables.

Variables level	$F1$ (t·h ⁻¹)	$F2$ (t·h ⁻¹)	$F3$ (t·h ⁻¹)	$F4$ (t·h ⁻¹)	Thermal load (kW)	Reflux (t·h ⁻¹)
Minimum	64.13	34.15	5.73	4.43	2.38	3.58
Maximum	90.62	52.35	19.41	12.72	3.95	8.88
Average	79.43	45.84	12.74	9.82	3.18	7.05
Variance	20.57	12.35	3.09	1.55	0.11	0.25
SD	4.54	3.51	1.76	1.24	0.33	0.50
-1	78.23	44.64	11.54	8.62	1.98	5.85
0	79.43	45.84	12.74	9.82	3.18	7.05
1	84.87	50.06	14.84	11.31	3.58	7.65

* $F1$, $F2$, $F3$ and $F4$ are tower feed rates (Fig. 1); SD is a standard deviation.

Table 2 Planning experiments with five variables.

Test	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>Rf</i>	ΔP^*	Freq. **	N° V***
1	-1	-1	-1	-1	-1	238.2	1	5
2	+1	-1	-1	-1	-1	239.1	19	5
3	-1	+1	-1	-1	-1	232.1	7	4
4	+1	+1	-1	-1	-1	236.0	4	5
5	-1	-1	+1	-1	-1	219.5	27	4
6	+1	-1	+1	-1	-1	239.0	1	5
7	-1	+1	+1	-1	-1	225.3	1	5
8	+1	+1	+1	-1	-1	289.1	1	5
9	-1	-1	-1	+1	-1	215.8	1	5
10	+1	-1	-1	+1	-1	239.0	1	5
11	-1	+1	-1	+1	-1	222.9	1	5
12	+1	+1	-1	+1	-1	231.6	6	5
13	-1	-1	+1	+1	-1	239.6	4	4
14	+1	-1	+1	+1	-1	229.5	2	5
15	-1	+1	+1	+1	-1	271.6	27	4
16	+1	+1	+1	+1	-1	299.8	9	5
17	-1	-1	-1	-1	+1	238.2	1	5
18	+1	-1	-1	-1	+1	240.3	5	5
19	-1	+1	-1	-1	+1	232.1	7	4
20	+1	+1	-1	-1	+1	236.0	4	5
21	-1	-1	+1	-1	+1	219.5	27	4
22	+1	-1	+1	-1	+1	241.7	41	4
23	-1	+1	+1	-1	+1	256.5	2	4
24	+1	+1	+1	-1	+1	289.1	1	5
25	-1	-1	-1	+1	+1	215.8	1	5
26	+1	-1	-1	+1	+1	241.2	13	4
27	-1	+1	-1	+1	+1	222.9	1	5
28	+1	+1	-1	+1	+1	231.6	6	5
29	-1	-1	+1	+1	+1	239.9	2	4
30	+1	-1	+1	+1	+1	229.5	22	5
31	-1	+1	+1	+1	+1	271.6	27	5
32	+1	+1	+1	+1	+1	299.8	9	5

F1*, *F2*, *F3* and *F4*—tower feed rates (Fig. 1); *kgf·cm⁻²; **frequency; *N° of variables.

using a database with a large number of points. This occurs because that industrial plants rarely work with variables over a broad operational range. Thus, the values of the variables must be solved by eliminating the operational variables to increase the probability of their occurrence.

A qualitative study was carried out using an empirical equation generated with six operational variables. It was predicted that this equation would

provide a good representation of the phenomenon because of the large number of non-existent points. When five variables were considered, the number of existent operational variables increased to 70% with respect to the evaluated data.

Using the same amplitude to determine the working variables and excluding the thermal load, the planning matrix was generated, as shown in Table 2.

The results of the qualitative study indicated which

variables affected the phenomenon. The data from Table 2 were adjusted to a plane with a determination coefficient (R^2) of 88% and a standard deviation of 11.4. The effect of each variable was calculated with 95% accuracy. These values are shown in Table 3 with their significance indicated.

Only the first three feed streams influenced the pressure variation in the column. Furthermore, only the first stream had significant interactions with the other variables. Thus, only the first feed stream influenced the pressure variation in the column independently of the values of the other variables.

3.2 First Empirical Model

A quantitative study was performed using the significant values from the planning matrix to determine the frequency that the set of variables appeared in the set of 10,000 points. This produced a set of existing points of approximately 95%, which provided a more representative model.

Initially, an attempt was made to adjust the data to a plane surface, but none of the variables were significant, with a coefficient of determination (R^2) of 60%. Thus, a star matrix was implemented to test the values from the central points. The data in this matrix were adjusted to a squared surface with the empirical function represented in Eq. (2). A coefficient of determination (R^2) of 63% and a standard deviation of 17.2 for the significant variables were calculated.

$$\begin{aligned} \Delta P = & 25228 + 1354F_1 + 1193F_2 + 1522F_3 - 5.97F_1^2 \\ & + 6.66F_2^2 - 3.64F_3^2 + 1.55F_1F_2 \\ & + 5.95F_1F_3 + 6.08F_2F_3 \end{aligned} \quad (2)$$

3.3 Validation of the First Empirical Model

Eq. (2) was validated with the 10,000 points from the database. The distribution of errors between the calculated and the real pressure variation is shown in Fig. 3.

The error distribution was generally negative for the low-feed streams and positive for the high-feed streams. In addition, Eq. (2) generated large errors

Table 3 Effects of independent variables.

Variable	ΔP (kgf·cm ⁻²)	Significant
F1 (1)	15.67	Yes
F2 (2)	20.14	Yes
F3 (3)	21.76	Yes
F4 (4)	1.90	No
Reflux (5)	2.35	No
(1)(2)	6.57	No
(1)(3)	6.08	No
(1)(4)	-2.94	No
(1)(5)	-1.59	No
(2)(3)	22.94	Yes
(2)(4)	5.05	No
(2)(5)	1.55	No
(3)(4)	10.80	Yes
(3)(5)	1.92	No
(4)(5)	-2.04	No

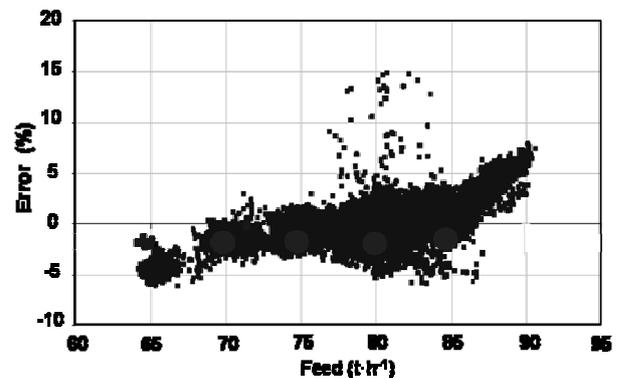


Fig. 3 Error distribution.

outside the central region and low sensitivity in the other regions. These tendencies indicated that Eq. (2) did not adequately represent the phenomenon.

3.4 Adaptation of the Model to the Real Industrial Plant

The need to determine a new equation led to the examination of data sets involving two and three variables, until a new equation was determined to adequately predict the maximum conditions of the plant. The effect of the total feed into the column was then examined while reducing the number of operational variables. This action influenced the ease at which points could be identified from the planning matrix. The planning matrix is shown in Table 4.

Table 4 Planning experiments.

Test	Ft	Rf	ΔP (kgf·cm ⁻²)
1	-1	-1	205.0
2	+1	-1	287.0
3	-1	+1	211.0
4	+1	+1	302.0
5	-1.41	0	228.4
6	1.41	0	235.1
7	0	-1.41	229.1
8	0	1.41	200.0
9	0	0	270.0
10	0	0	218.0
11	0	0	228.0

The data from Table 4 were adjusted to a squaredplane, generating a coefficient of determination of 99% and a standard deviation of 6.8. Eq. (3) was the empirical function that represented the phenomenon.

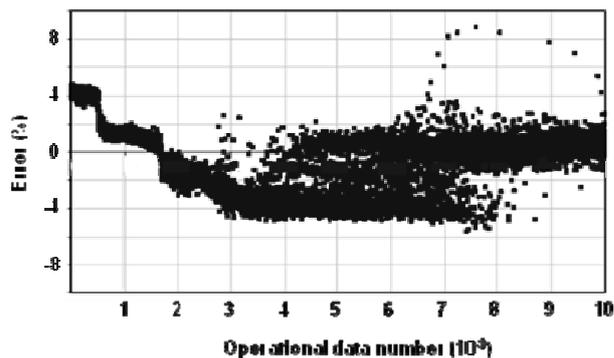
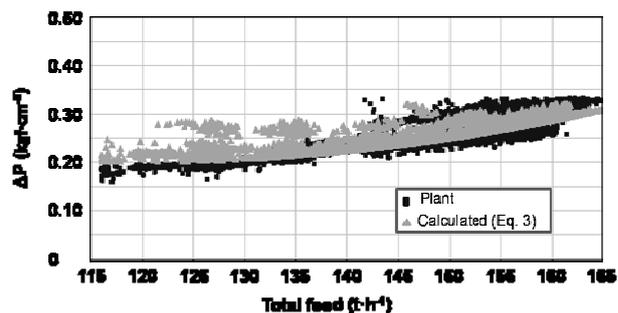
$$\Delta P = 230.87 + 42.48 Ft + 10.05 Rf + 13.38 Ft^2 + 8.38 Rf^2 + 2.25 Ft.Rf \quad (3)$$

Eq. (3) was tested using the same 10,000 points from the database. The responses showed that the model properly represented all of the studied regions. However, Fig. 4 shows that the errors generated by Eq. (3) together with their distributions depend on the amount of operational data collected from the plant.

Fig. 4 can be divided into three zones. Zone I includes the errors of the first 4,000 points. Zone II includes the errors between points 4,000 and 8,000. Zone III includes the errors found from point 8,000 and up. The best results were found in Zone III, where the error distribution was random and the values were within $\pm 2\%$. Thus, Eq. (3) was well suited to evaluate the data in this step.

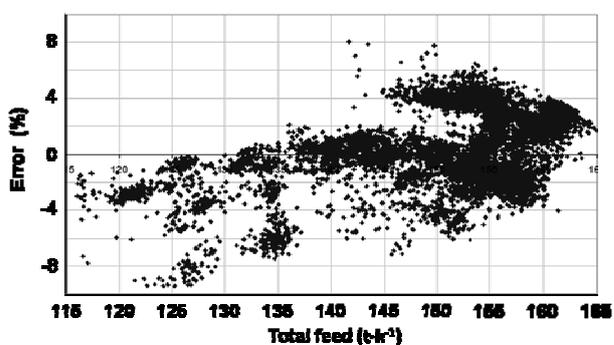
The robustness of Eq. (3) was assessed using a new set of operational data. The set was formed from 10,000 new points collected months after the first data set. The results are shown in Fig. 5.

Eq. (3) generated values for pressure variation close to those obtained for the real operation of the demethanizer tower. The prediction was more efficient for high total flow rates but adequately represented


Fig. 4 Error distribution.

Fig. 5 Evaluation of the empirical model.

the process over the entire range. Fig. 6 shows that the majority of errors were found within the $\pm 4\%$ range. In addition, the errors being were best distributed for flow rates higher than $135 \text{ t}\cdot\text{h}^{-1}$.

Eq. (3) generated values for pressure variation close to those obtained for the real operation of the demethanizer tower. The prediction was more efficient for high total flow rates but adequately represented the process over the entire range. Fig. 6 shows that the majority of errors were found within the $\pm 4\%$ range. In addition, the errors being were best distributed for flow rates higher than $135 \text{ t}\cdot\text{h}^{-1}$.


Fig. 6 Error distribution.

4. Conclusions

Eq. (3) was used in the simulations to predict the feed rate and reflux rate limits of the demethanizer tower with no flooding. The first step of the calculation was to determine the feed rate and reflux rate of the column using a commercial simulator based on the load to be processed in the plant. Next, the data calculated by the commercial simulator were applied to the model (Eq. (3)) to prevent flooding. The model identified the presence of flooding from the calculation of the pressure variation in the tower, so a new operational condition was determined considering the maximum flow rate that would not cause flooding. The column feed excess was diverted and processed in another plant. This procedure eliminated losses caused by flooding and brought about gains because it allowed processing of greater loads in the malfunctioning demethanizer.

Eq. (3) was used to predict the maximum column load for approximately 20 months until the equipment was stopped for maintenance, at which time the last column tray was changed.

In conclusion, this methodology will be useful for predicting equipment operation, especially when phenomenological equations cannot represent the malfunctioning piece of equipment, which is a very common event in industry.

References

- [1] Kresta, J., MacGregor, J. F., and Marlin, T. E. 1991. "Multivariate Statistical Monitoring of Process Operating Performance." *Can. J. Chem. Eng.* 69: 35-47.
- [2] Kourti, T. 2006. "The Process Analytical Technology Initiative and Multivariate Process Analysis, Monitoring and Control." *Analyt. Bioanalyt. Chem.* 384 (5): 1043-8.
- [3] Kano, M., Hashimoto, I., and Ohno, H. 2004. "Evolution of Multivariate Statistical Process Control: Application of Independent Component Analysis and External Analysis." *Comput. Chem. Eng.* 28: 1157-66.
- [4] Martin, E.B., Morris, A.J., and Zhang, J. 1996. "Process Performance Monitoring using Multivariate Process Control." *IEE Proc.-Control Theory Appl.* 143: 132-44.
- [5] Zhang, J., Martin, E.B., and Morris, A.J. 1997. "Process Monitoring Using Non-Linear Statistical Techniques." *Chem. Eng. J.* 67: 181-9.
- [6] Kleijnen, J. P. C. 1990. "Statistics and Deterministic Simulation Models: Why Not?" In *Winter Simulation Conference Proceedings of the 22nd Conference on Winter Simulation*, 344-6.
- [7] Zullo, L. 1996. "Validation and Verification of Continuous Plants Operating Modes Using Multivariate Statistical Methods." *Comput. Chem. Eng.* 20: 5683-8.
- [8] Kourti, T., Lee, J., and Macgregor, J. F. 1996. "Experiences with Industrial Applications of Projection Methods for Multivariate Statistical Process Control." *Comput. Chem. Eng.* 20: 5745-50.
- [9] Kano, M., Hashimoto, I., Ohno, H., and Hasebe, S. 2001. "A New Multivariate Statistical Process Monitoring Method Using Principal Component Analysis." *Comput. Chem. Eng.* 25: 1103-13.
- [10] Lee, J. M., Yoo, C., and Lee, I. B. 2004. "Statistical Monitoring of Dynamic Process Based on Dynamic Independent Component Analysis." *Chem. Eng. Sci.* 59: 223-34.
- [11] Soave, G. S., Gamba, S., Pellegrini, L. A., and Bonomi, S. 2006. "Feed-Splitting Technique in Cryogenic Distillation." *Ind. Eng. Chem. Res.* 45: 5761-5.